
CyVerse Data Commons: lessons learned in cyberinfrastructure management and data hosting from the Life Sciences
AGU Fall Meeting December 11, 2017 Sesson IN12B-07
Tyson L. Swetnam, Ramona L. Walls, & Nirav Merchant
BIO5 Institute, University of Arizona, Tucson AZ
Enter Here
CYVERSE


Overview
About
Data Store & Data Commons
Discovery Environment
Atmosphere
Science APIs
Lessons Learned
Considerations for choosing the appropriate data reposito...
Verse = Universe
Data Store
Monitor job status, View results
Atmosphere
Science APIs
Lessons learned in cyberinfrastructure management and data hosting
Acknowledgements

Considerations for choosing the appropriate data repository for your science:

1) Are the data manageable throughout their entire life cycle?

2) Is there user support for the data in the repository?

3) Are the data searchable (indexed), accessible, for computation?


CYVERSE®


CYVERSE®
Cy = Cyber


CYVERSE®
Cy = Cyber
Verse = Universe

CYVERSE®

Cy = Cyber Verse = Universe

Vision

Mission

Usage

CYVERSE®

Vision

Mission

Usage

CYVERSE®

Vision

Design, develop, deploy, and expand a national cyberinfrastructure for life science research, and train scientists in its use.

Mission

Usage

CYVERSE®

Vision Design, develop, deploy, and expand a national cyberinfrastructure for life science research, and train scientists in its use.

Mission Transforming science through data-driven discovery

Usage

CYVERSE®

Vision Design, develop, deploy, and expand a national cyberinfrastructure for life science research, and train scientists in its use.

Mission Transforming science through data-driven discovery

Usage 10th year, >50K Users, 3+ PB data, 850+ peer pubs, hosted 100s workshops, teaching courses

CYVERSE®

Vision Design, develop, deploy, and expand a national cyberinfrastructure for life science research, and train scientists in its use.

Biogeosciences = Life Science

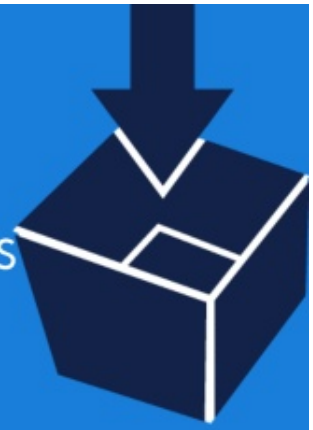
Mission Transforming science through data-driven discovery

Usage 10th year, >50K Users, 3+ PB data, 850+ peer pubs, hosted 100s workshops, teaching courses


Ease of Use
Flexibility
PRODUCTS
Atmosphere
Discovery Environment
DNA Subway
BisQue Image Analysis
Powered by CYVERSE
Third Party Consumers
Data Commons
Ready to use Platforms
FOUNDATIONAL SERVICES
Data Store
Science APIs
Auth
elastic
Extensible Services
LOW-LEVEL SERVICES, SECURITY, ASSETS, etc.
iRODS (federated storage)
CAS/Shibboleth/OAuth2 (single sign-on)
OpenStack (virtualization)
Condor (job scheduling)
XSEDE (national CI)
Established CI Components
HARDWARE RESOURCES
Cloud Systems
High-performance Computers
Databases
Storage
Foundational Capabilities

Data Store

- 2.8 PB + 50TB/month
- 72 Mil Files
- 236 Mil Metadata elements
- 8,620 users



Data Store

- 2.8 PB + 50TB/month
- 72 Mil Files
- 236 Mil Metadata elements
- 8,620 users

- High Performance cloud storage and management

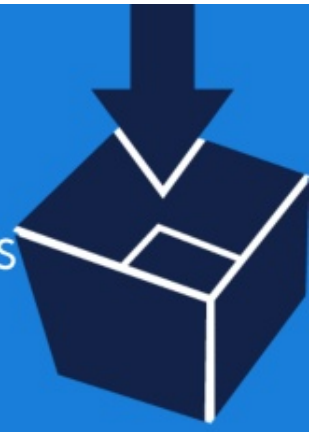
Data Store

- 2.8 PB + 50TB/month
- 72 Mil Files
- 236 Mil Metadata elements
- 8,620 users

- High Performance cloud storage and management
- Multi-threaded file transfer (iRODS iCommands)

iRODS

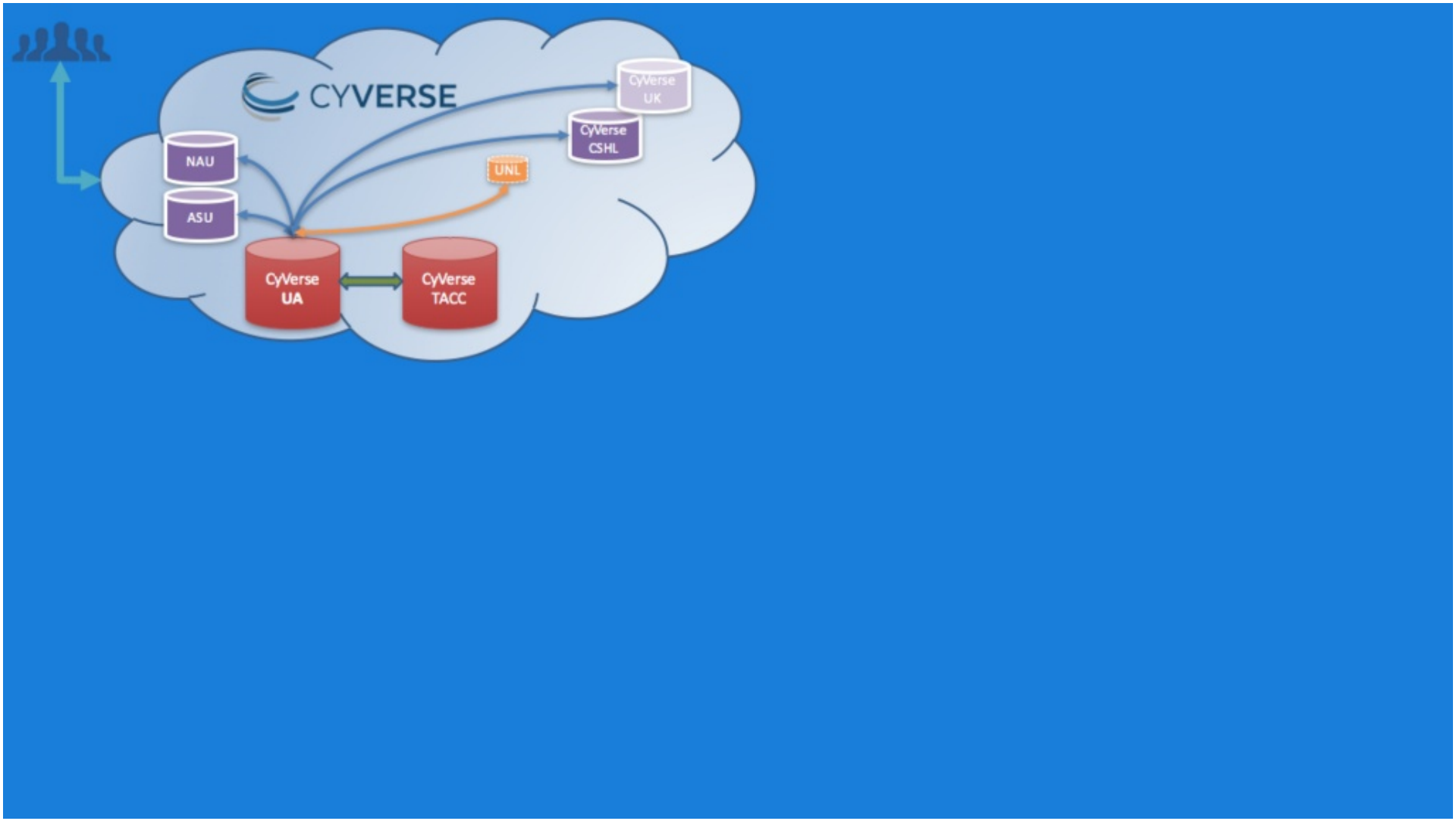
Data Store

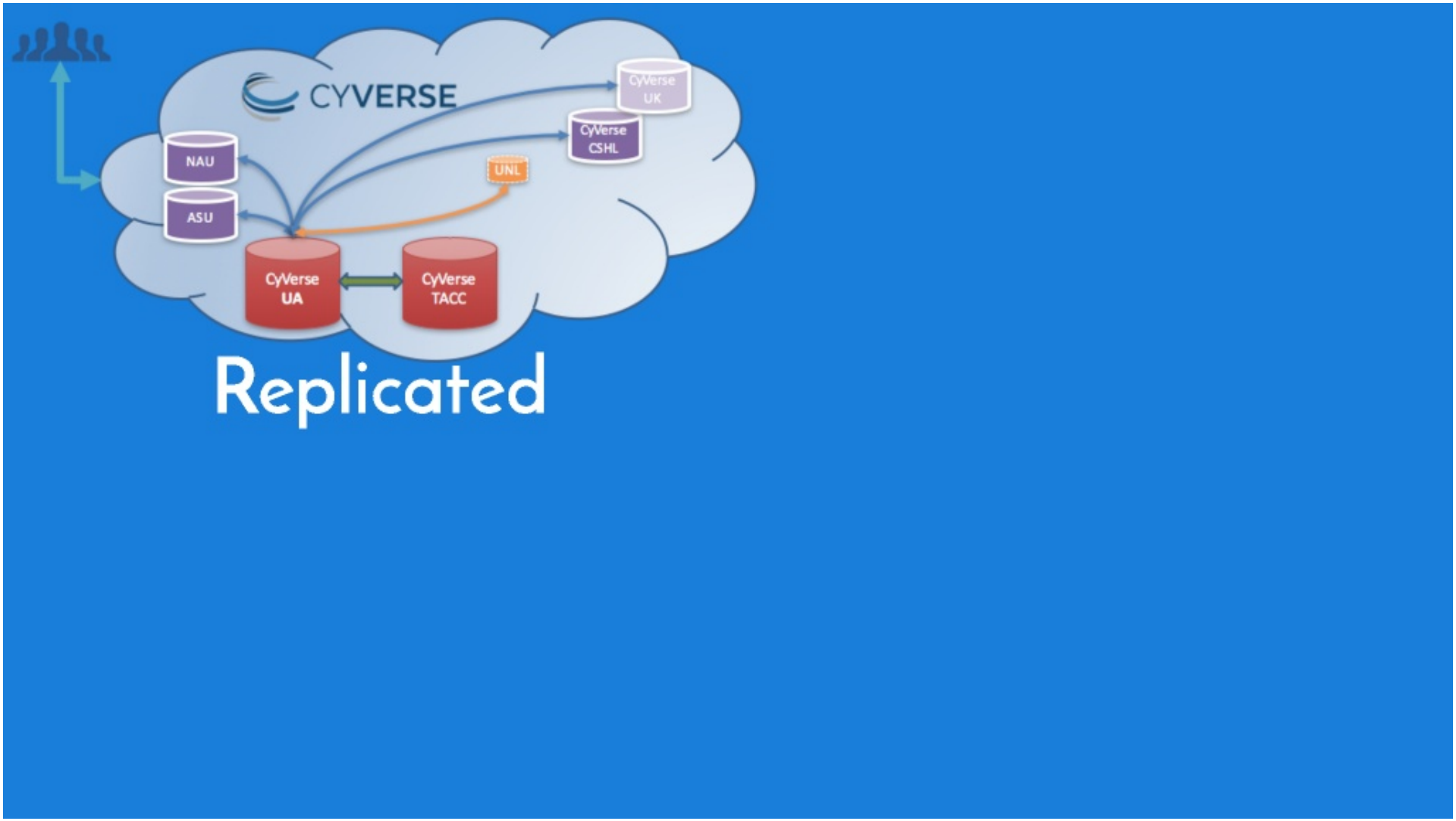


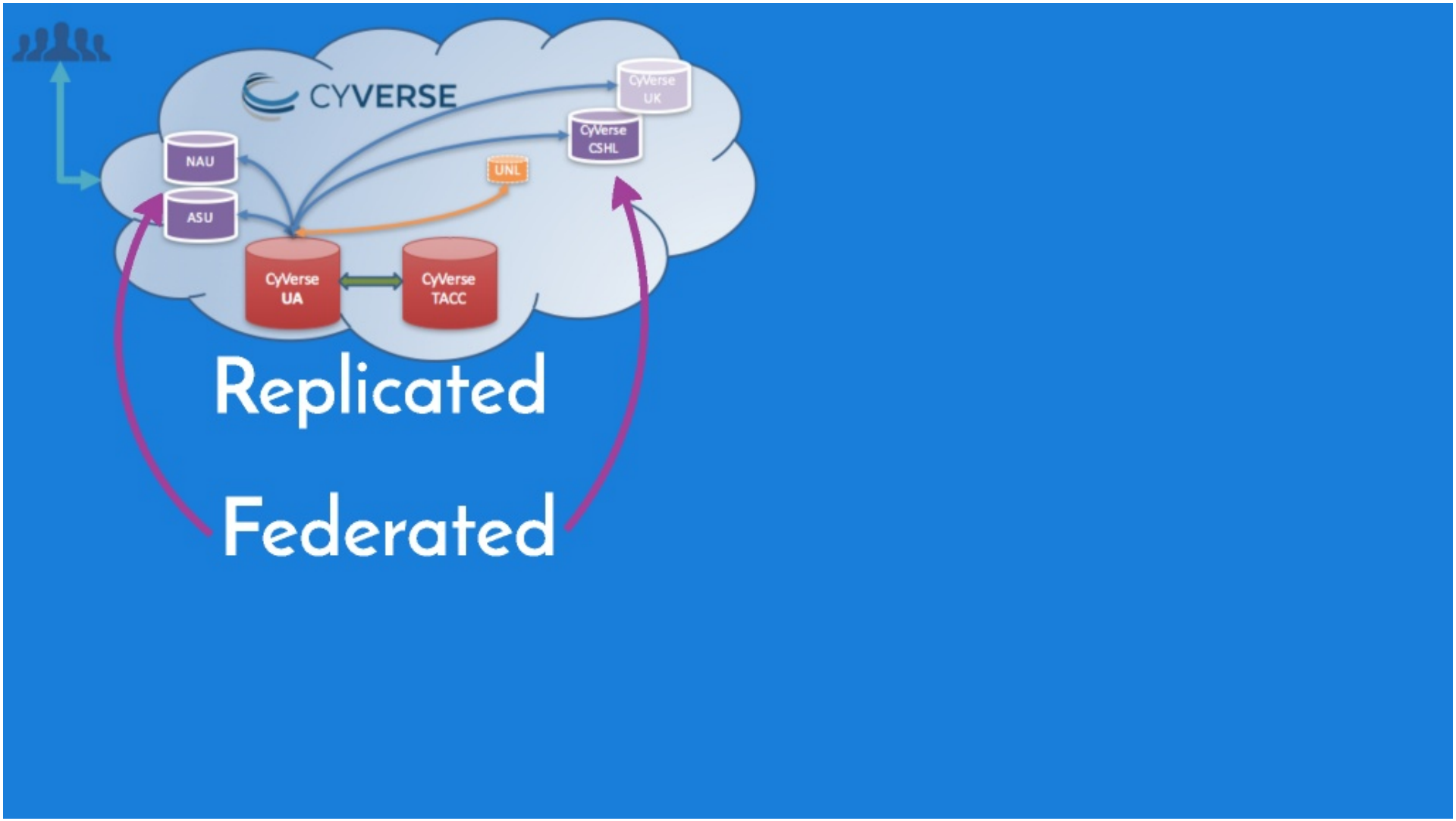
- 2.8 PB + 50TB/month
- 72 Mil Files
- 236 Mil Metadata elements
- 8,620 users

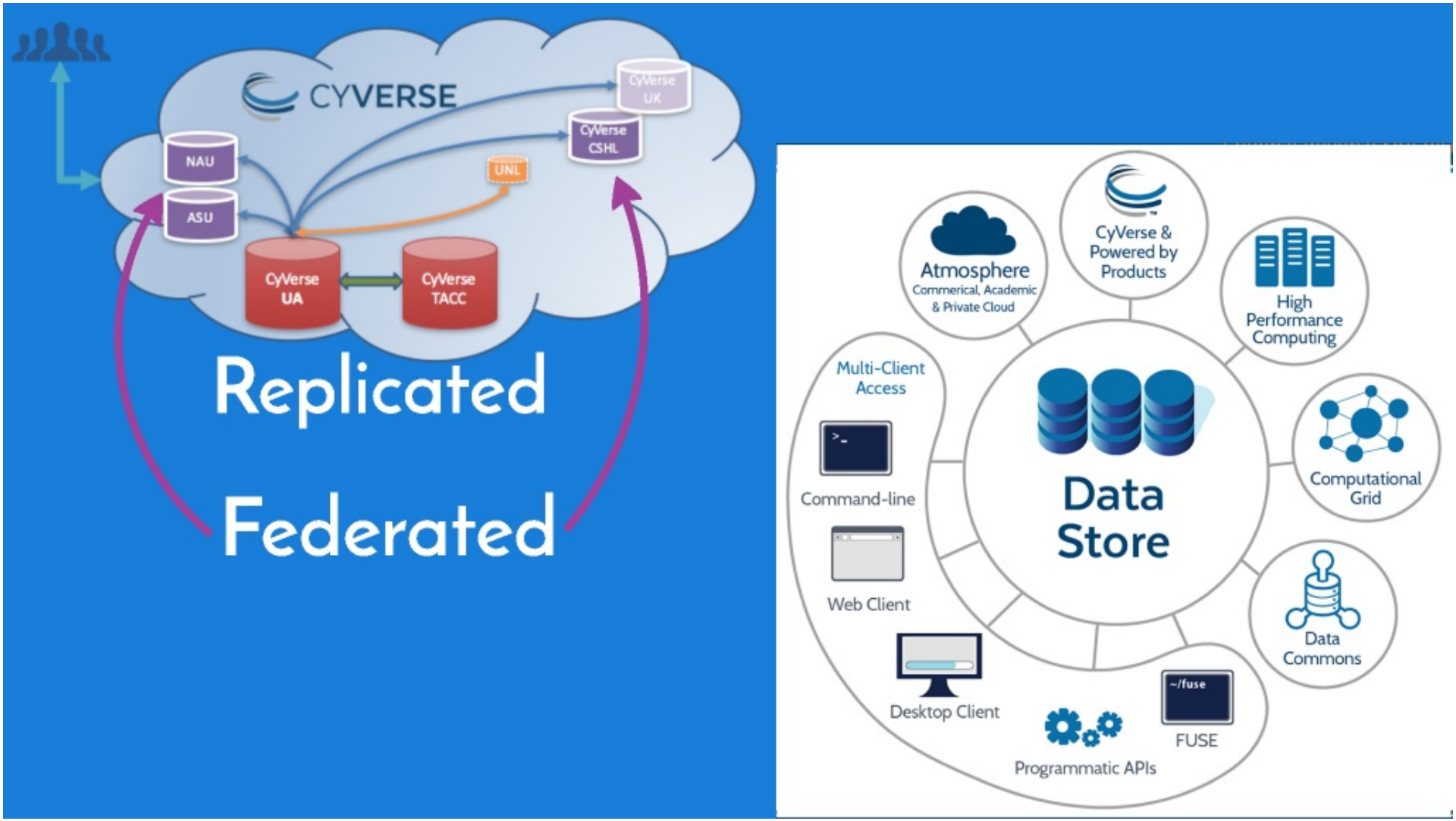
- High Performance cloud storage and management
- Multi-threaded file transfer (iRODS iCommands)
- Interface in Python, C, Java, and REST API (CyberDuck)

iRODS

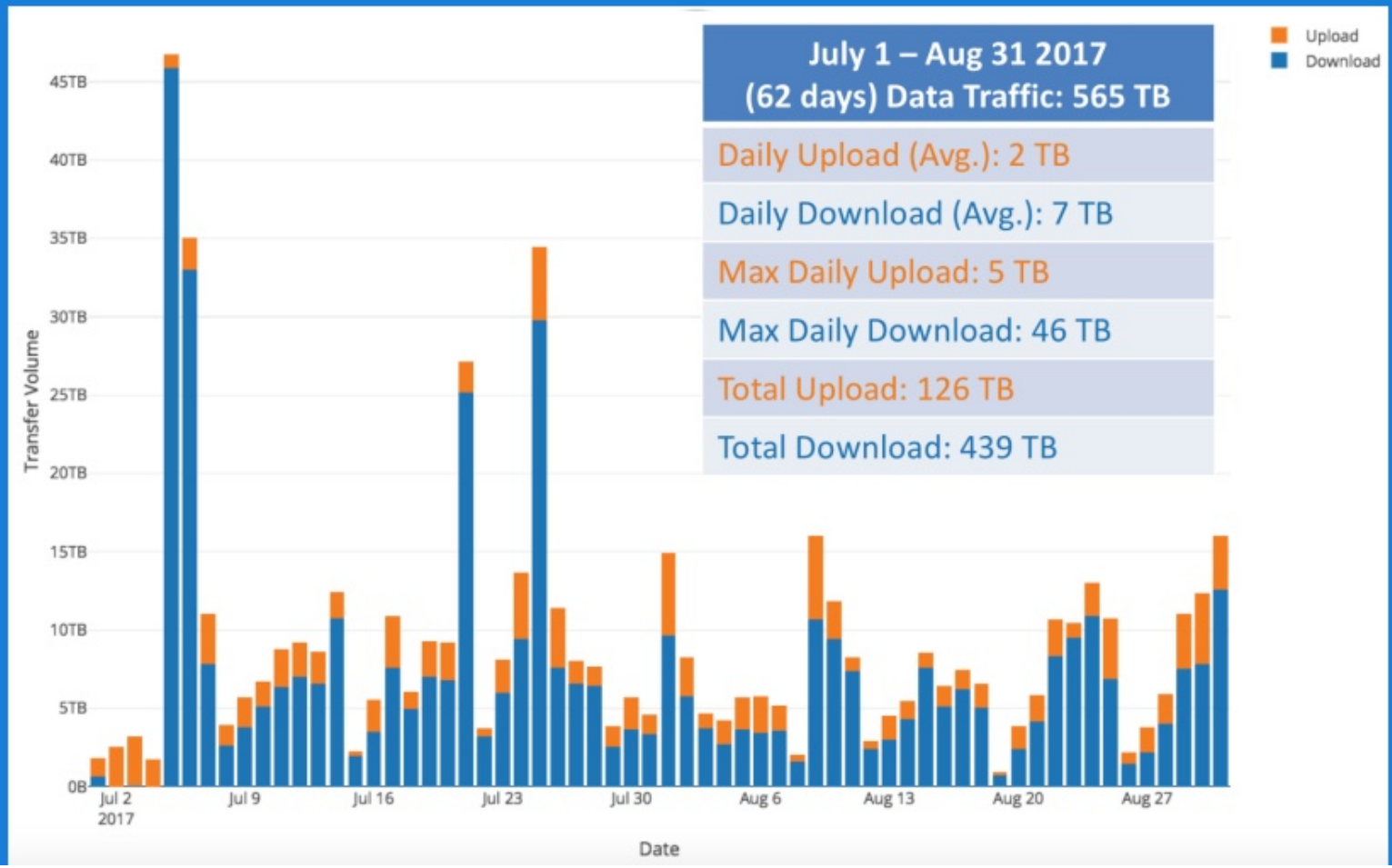

CYVERSE
CyVerse UK
CyVerse CSHL
NAU
UNL
ASU
CyVerse UA
CyVerse TACC


CYVERSE
CyVerse UK
CyVerse CSHL
NAU
UNL
ASU
CyVerse UA
CyVerse TACC
Replicated

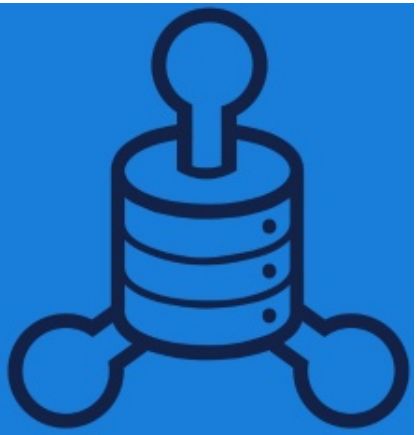

CYVERSE
CyVerse UK
CyVerse CSHL
NAU
UNL
ASU
CyVerse UA
CyVerse TACC
Replicated
Federated


CYVERSE
CyVerse UK
CyVerse CSHL
NAU
UNL
ASU
CyVerse UA
CyVerse TACC
Replicated
Federated
Atmosphere
Commerical, Academic & Private Cloud
CyVerse & Powered by Products
High Performance Computing
Multi-Client Access
Command-line
Web Client
Data Store
Computational Grid
Data Commons
Desktop Client
~/fuse
FUSE
Programmatic APIs

Data Store


July 1 – Aug 31 2017
(62 days) Data Traffic: 565 TB
Daily Upload (Avg.): 2 TB
Daily Download (Avg.): 7 TB
Max Daily Upload: 5 TB
Max Daily Download: 46 TB
Total Upload: 126 TB
Total Download: 439 TB
Upload
Download
Transfer Volume
45TB
40TB
35TB
30TB
25TB
20TB
15TB
10TB
5TB
0B
Jul 2
2017
Jul 9
Jul 16
Jul 23
Jul 30
Aug 6
Aug 13
Aug 20
Aug 27
Date

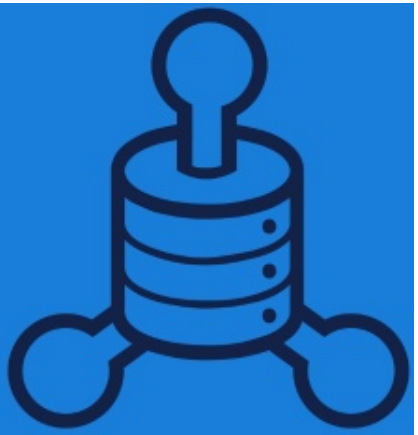
Data Commons



http://datacommons.cyverse.org/

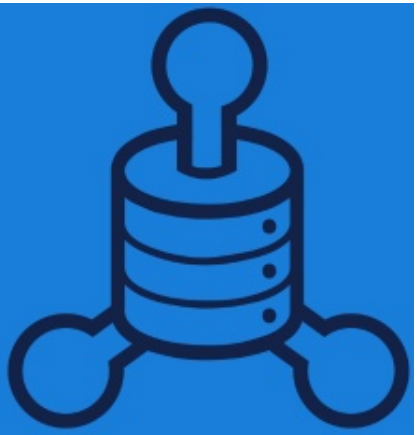
Data Commons

Champion development of CyVerse features for researchers to organize, preserve, and publish data derived from scientific research.



http://datacommons.cyverse.org/

Data Commons

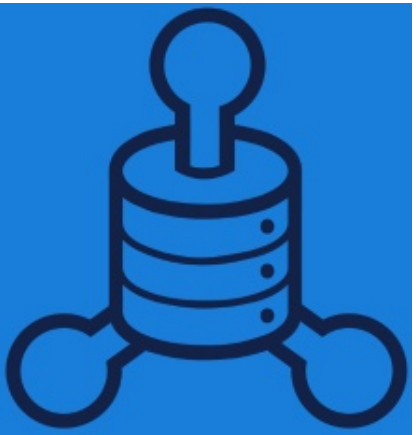


Champion development of CyVerse features for researchers to organize, preserve, and publish data derived from scientific research.

- Metadata / Ontologies
- Permanant Identifier
- Community Standards
- Data discovery

http://datacommons.cyverse.org/

Data Commons



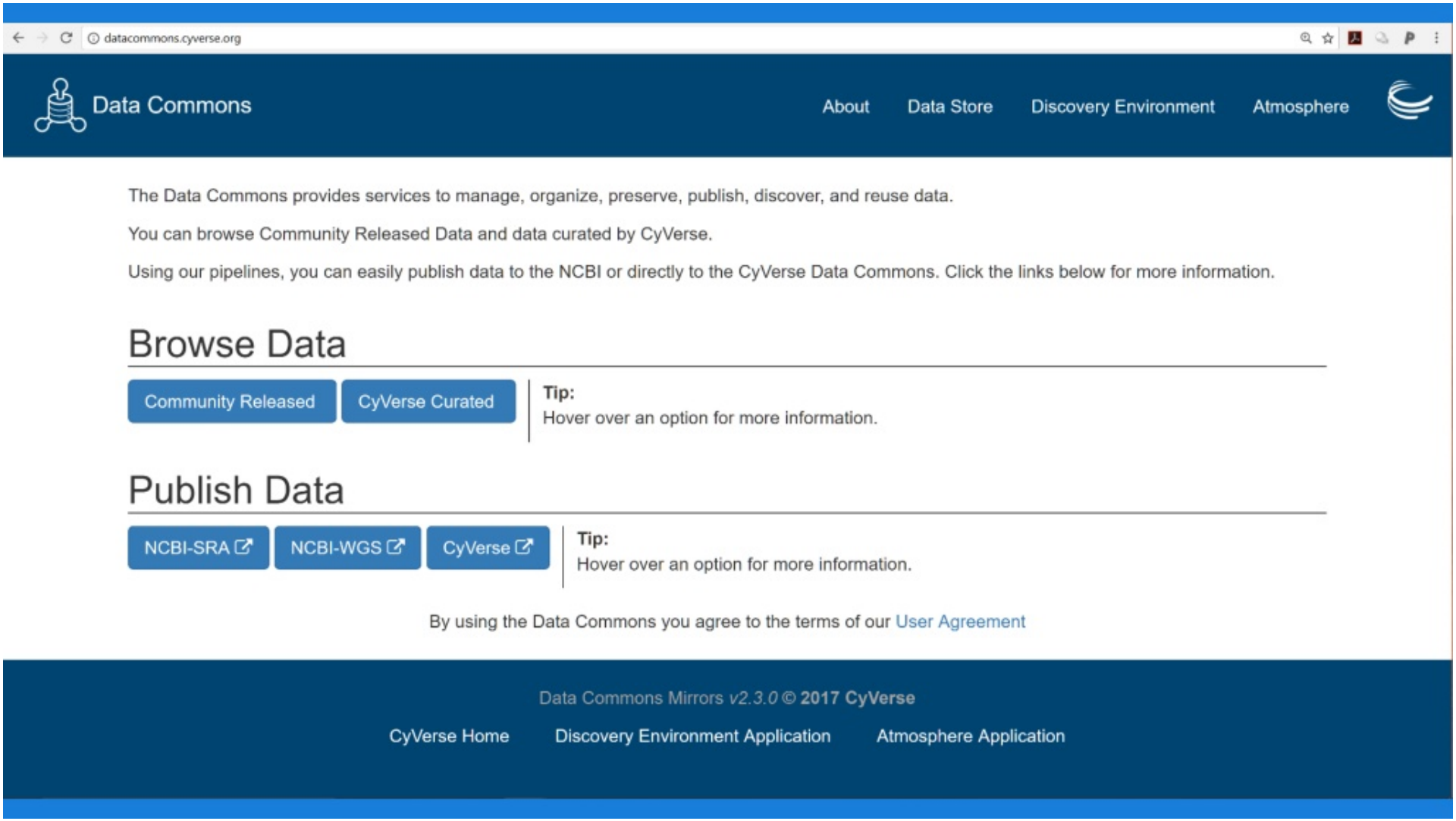
Champion development of CyVerse features for researchers to organize, preserve, and publish data derived from scientific research.

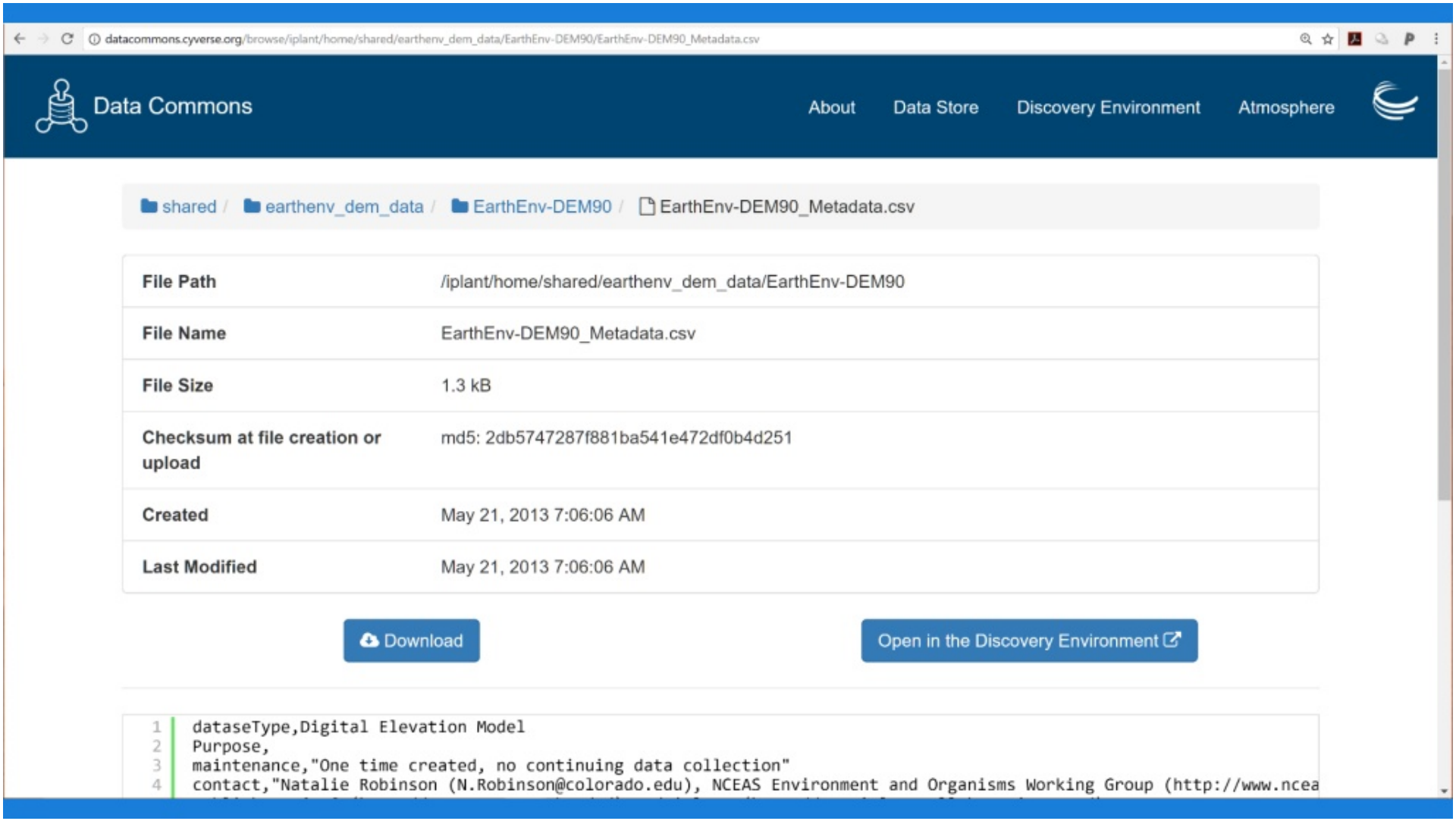
- Metadata / Ontologies
- Permanant Identifier
- Community Standards
- Data discovery

Coming Soon:
Member Node

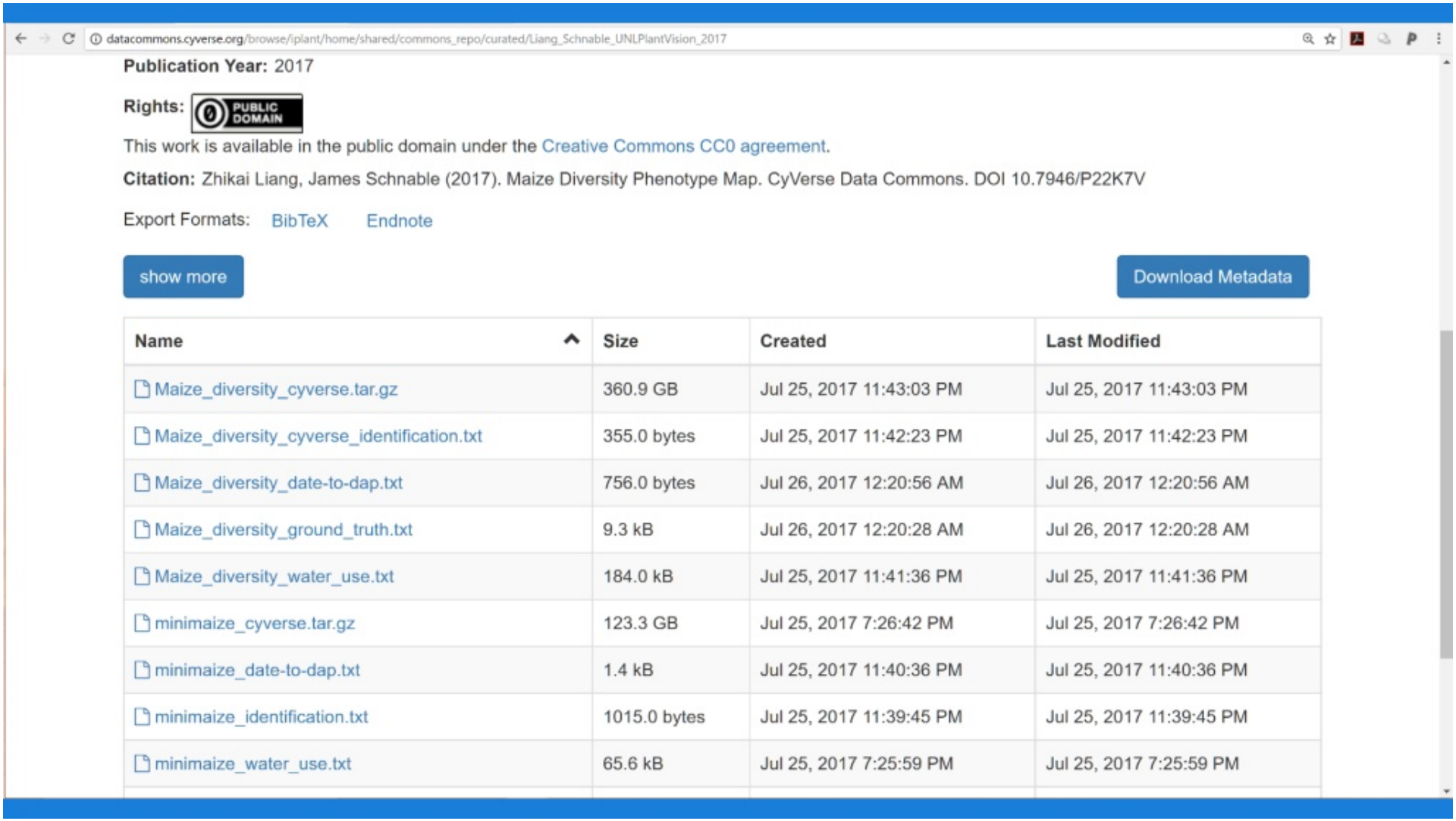
DataONE
Data Observation Network for Earth

http://datacommons.cyverse.org/


datacommons.cyverse.org
Data Commons
About
Data Store
Discovery Environment
Atmosphere
The Data Commons provides services to manage, organize, preserve, publish, discover, and reuse data.
You can browse Community Released Data and data curated by CyVerse.
Using our pipelines, you can easily publish data to the NCBI or directly to the CyVerse Data Commons. Click the links below for more information.
Browse Data
Community Released
CyVerse Curated
Tip:
Hover over an option for more information.
Publish Data
NCBI-SRA
NCBI-WGS
CyVerse
Tip:
Hover over an option for more information.
By using the Data Commons you agree to the terms of our User Agreement
Data Commons Mirrors v2.3.0 © 2017 CyVerse
CyVerse Home
Discovery Environment Application
Atmosphere Application

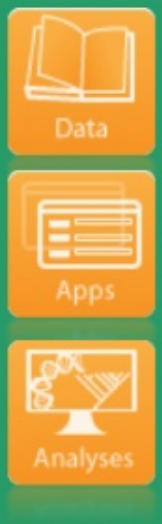

datacommons.cyverse.org/browse/iplant/home/shared/earthenv_dem_data/EarthEnv-DEM90/EarthEnv-DEM90_Metadata.csv
Data Commons
About
Data Store
Discovery Environment
Atmosphere
shared / earthenv_dem_data / EarthEnv-DEM90 / EarthEnv-DEM90_Metadata.csv
File Path
/iplant/home/shared/earthenv_dem_data/EarthEnv-DEM90
File Name
EarthEnv-DEM90_Metadata.csv
File Size
1.3 kB
Checksum at file creation or upload
md5: 2db5747287f881ba541e472df0b4d251
Created
May 21, 2013 7:06:06 AM
Last Modified
May 21, 2013 7:06:06 AM
Download
Open in the Discovery Environment
1 dataseType,Digital Elevation Model
2 Purpose,
3 maintenance,"One time created, no continuing data collection"
4 contact,"Natalie Robinson (N.Robinson@colorado.edu), NCEAS Environment and Organisms Working Group (http://www.ncea


datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Liang_Schnable_UNLPlantVision_2017
Data Commons
About
Data Store
Discovery Environment
Atmosphere
shared / commons_repo / curated / Liang_Schnable_UNLPlantVision_2017
Maize Diversity Phenotype Map
DOI: 10.7946/P22K7V
Creator: Zhikai Liang, James Schnable
Description: This directory includes image data collected from 155 maize plants over 32 days representing 32 maize inbreds which were also grown across the USA for field phenotyping as part the of the Genomes to Fields (G2F) project. Plants were grown and data collected at the University of Nebraska Automated Greenhouse facility in Fall 2015. Each imaging time point included data from four different types of cameras: RGB, Hyperspectral, Fluorescent, and Thermal IR. Plant water consumption was digitally recorded each day using changes in pot weight. The genotypes imaged here are linked with ground truth data collected from the same plants, field phenotyping data from the same plant varieties grown in multiple states in multiple years (https://doi.org/10.7946/P2V888, https://doi.org/10.7946/P24S31), and GBS-based genotype calls (G2F: https://doi.org/10.7946/P2V888, Romay MC et al.(2013): DOI: 10.1186/gb-2013-14-6-r55). Because plant phenomics is an emerging area, it is hoped that the prepublication release of these novel data types will encourage the development of new image analysis methods in the emerging field of plant phenomics as well as new models for connecting phenotypic data across multiple environments.
Publisher: CyVerse Data Commons
Publication Year: 2017
Rights:
PUBLIC DOMAIN
This work is available in the public domain under the Creative Commons CC0 agreement.
Citation: Zhikai Liang, James Schnable (2017). Maize Diversity Phenotype Map. CyVerse Data Commons. DOI 10.7946/P22K7V


datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Liang_Schnable_UNLPlantVision_2017
Publication Year: 2017
Rights:
PUBLIC DOMAIN
This work is available in the public domain under the Creative Commons CC0 agreement.
Citation: Zhikai Liang, James Schnable (2017). Maize Diversity Phenotype Map. CyVerse Data Commons. DOI 10.7946/P22K7V
Export Formats: BibTeX Endnote
show more
Download Metadata
Name
Size
Created
Last Modified
Maize_diversity_cyverse.tar.gz 360.9 GB Jul 25, 2017 11:43:03 PM Jul 25, 2017 11:43:03 PM
Maize_diversity_cyverse_identification.txt 355.0 bytes Jul 25, 2017 11:42:23 PM Jul 25, 2017 11:42:23 PM
Maize_diversity_date-to-dap.txt 756.0 bytes Jul 26, 2017 12:20:56 AM Jul 26, 2017 12:20:56 AM
Maize_diversity_ground_truth.txt 9.3 kB Jul 26, 2017 12:20:28 AM Jul 26, 2017 12:20:28 AM
Maize_diversity_water_use.txt 184.0 kB Jul 25, 2017 11:41:36 PM Jul 25, 2017 11:41:36 PM
minimaize_cyverse.tar.gz 123.3 GB Jul 25, 2017 7:26:42 PM Jul 25, 2017 7:26:42 PM
minimaize_date-to-dap.txt 1.4 kB Jul 25, 2017 11:40:36 PM Jul 25, 2017 11:40:36 PM
minimaize_identification.txt 1015.0 bytes Jul 25, 2017 11:39:45 PM Jul 25, 2017 11:39:45 PM
minimaize_water_use.txt 65.6 kB Jul 25, 2017 7:25:59 PM Jul 25, 2017 7:25:59 PM

Discovery Environment

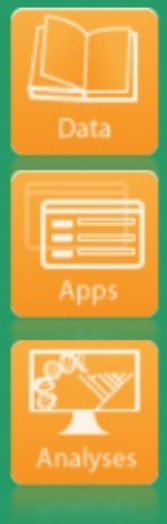
3,970 Apps, 259,219 analysis, 5,740 users



http://de.cyverse.org

Discovery Environment

3,970 Apps, 259,219 analysis, 5,740 users



Find a file or upload new data to the Data Store

http://de.cyverse.org

Discovery Environment

3,970 Apps, 259,219 analysis, 5,740 users

Find a file or upload new data to the Data Store

Use or Build Tools with Docker Containers

http://de.cyverse.org

Discovery Environment

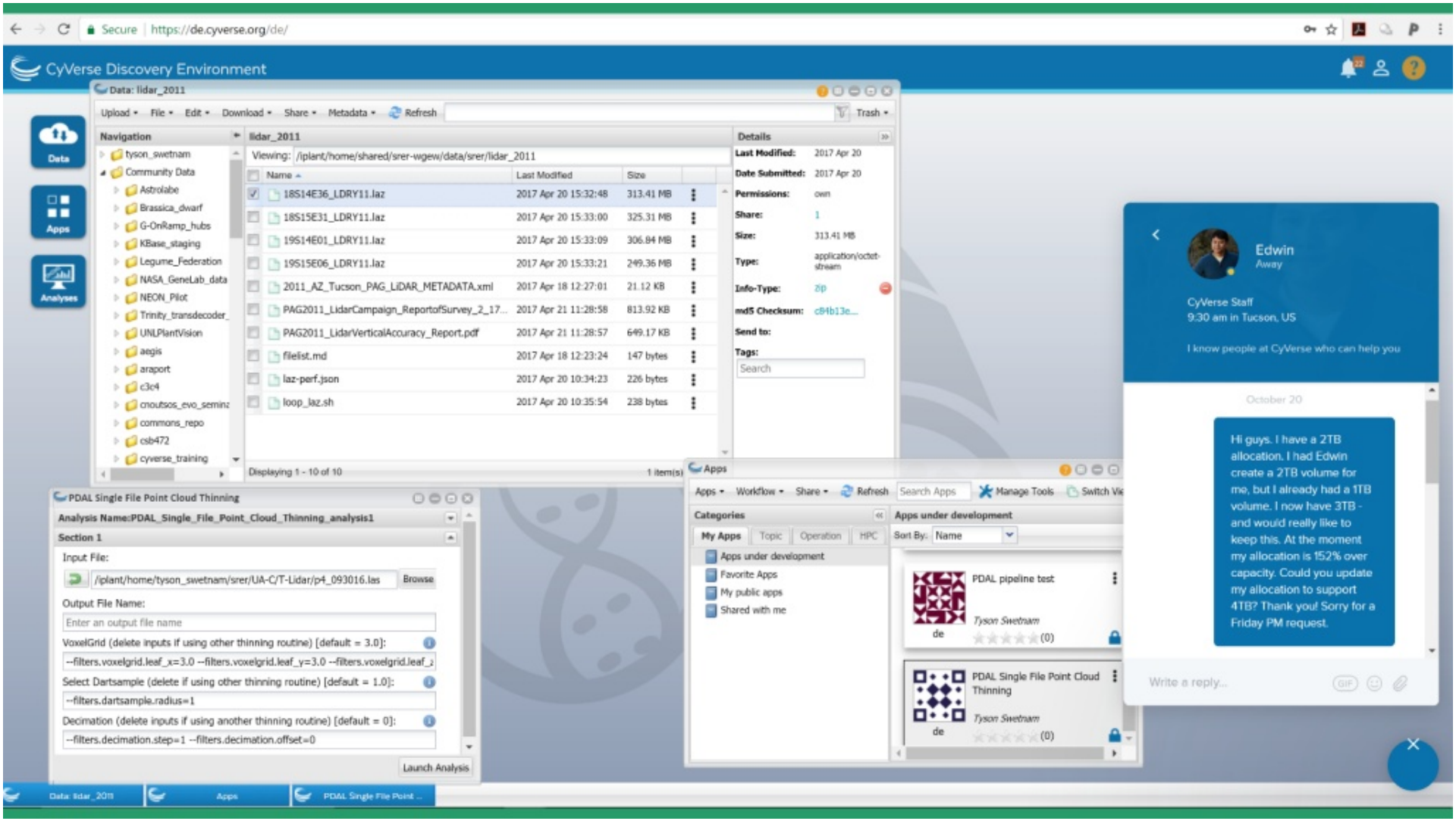
3,970 Apps, 259,219 analysis, 5,740 users

Find a file or upload new data to the Data Store

Use or Build Tools with Docker Containers

Monitor job status, View results

http://de.cyverse.org


Secure | https://de.cyverse.org/de/
CyVerse Discovery Environment
Data: lidar_2011
Upload File Edit Download Share Metadata Refresh Trash
Navigation
Data
Apps
Analyses
tyson_swetnam
Community Data
Astrolabe
Brassica_dwarf
G-OnRamp_hubs
KBase_staging
Legume_Federation
NASA_GeneLab_data
NEON_Pilot
Trinity_transdecoder_
UNLPlantVision
aegis
araport
c3c4
cnoutsos_evo_semina
commons_repo
csb472
cyverse_training
lidar_2011
Viewing: /iplant/home/shared/srer-wgew/data/srer/lidar_2011
Name
Last Modified
Size
18S14E36_LDRY11.laz 2017 Apr 20 15:32:48 313.41 MB
18S15E31_LDRY11.laz 2017 Apr 20 15:33:00 325.31 MB
19S14E01_LDRY11.laz 2017 Apr 20 15:33:09 306.84 MB
19S15E06_LDRY11.laz 2017 Apr 20 15:33:21 249.36 MB
2011_AZ_Tucson_PAG_LiDAR_METADATA.xml 2017 Apr 18 12:27:01 21.12 KB
PAG2011_LidarCampaign_ReportofSurvey_2_17... 2017 Apr 21 11:28:58 813.92 KB
PAG2011_LidarVerticalAccuracy_Report.pdf 2017 Apr 21 11:28:57 649.17 KB
filelist.md 2017 Apr 18 12:23:24 147 bytes
laz-perf.json 2017 Apr 20 10:34:23 226 bytes
loop_laz.sh 2017 Apr 20 10:35:54 238 bytes
Displaying 1 - 10 of 10
1 item(s)
Details
Last Modified: 2017 Apr 20
Date Submitted: 2017 Apr 20
Permissions: own
Share: 1
Size: 313.41 MB
Type: application/octet-stream
Info-Type: zip
md5 Checksum: c84b13e...
Send to:
Tags:
Search
PDAL Single File Point Cloud Thinning
Analysis Name:PDAL_Single_File_Point_Cloud_Thinning_analysis1
Section 1
Input File:
/iplant/home/tyson_swetnam/srer/UA-C/T-Lidar/p4_093016.las
Browse
Output File Name:
Enter an output file name
VoxelGrid (delete inputs if using other thinning routine) [default = 3.0]:
--filters.voxelgrid.leaf_x=3.0 --filters.voxelgrid.leaf_y=3.0 --filters.voxelgrid.leaf_z
Select Dartsample (delete if using other thinning routine) [default = 1.0]:
--filters.dartsample.radius=1
Decimation (delete inputs if using another thinning routine) [default = 0]:
--filters.decimation.step=1 --filters.decimation.offset=0
Launch Analysis
Apps
Apps Workflow Share Refresh Search Apps Manage Tools Switch Vie
Categories
My Apps Topic Operation HPC
Apps under development
Favorite Apps
My public apps
Shared with me
Apps under development
Sort By: Name
PDAL pipeline test
Tyson Swetnam
de
(0)
PDAL Single File Point Cloud Thinning
Tyson Swetnam
de
(0)
Edwin
Away
CyVerse Staff
9:30 am in Tucson, US
I know people at CyVerse who can help you
October 20
Hi guys. I have a 2TB allocation. I had Edwin create a 2TB volume for me, but I already had a 1TB volume. I now have 3TB - and would really like to keep this. At the moment my allocation is 152% over capacity. Could you update my allocation to support 4TB? Thank you! Sorry for a Friday PM request.
Write a reply...
Data: lidar_2011
Apps
PDAL Single File Point ...

Discovery Environment

- Docker Hub is the repository for DE docker containers
- If you create a workflow using multiple containers, this can be published in the Data Commons with a DOI
- (Coming Soon) Interactive Jobs running Jupyter Notebooks

Atmosphere

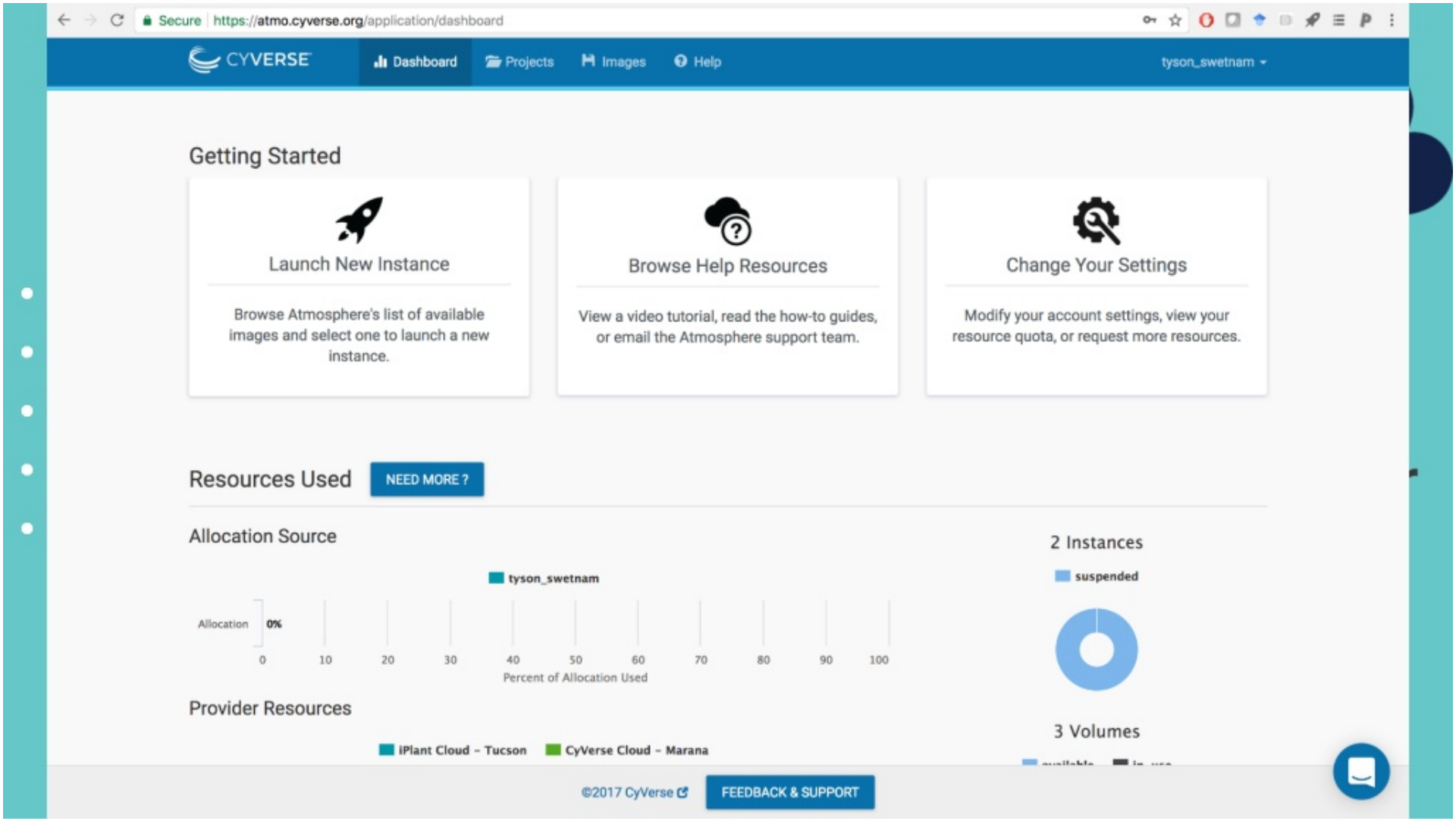
- On demand
- Linux environment
- Collaborate
- Publish images
- Easy to use interface

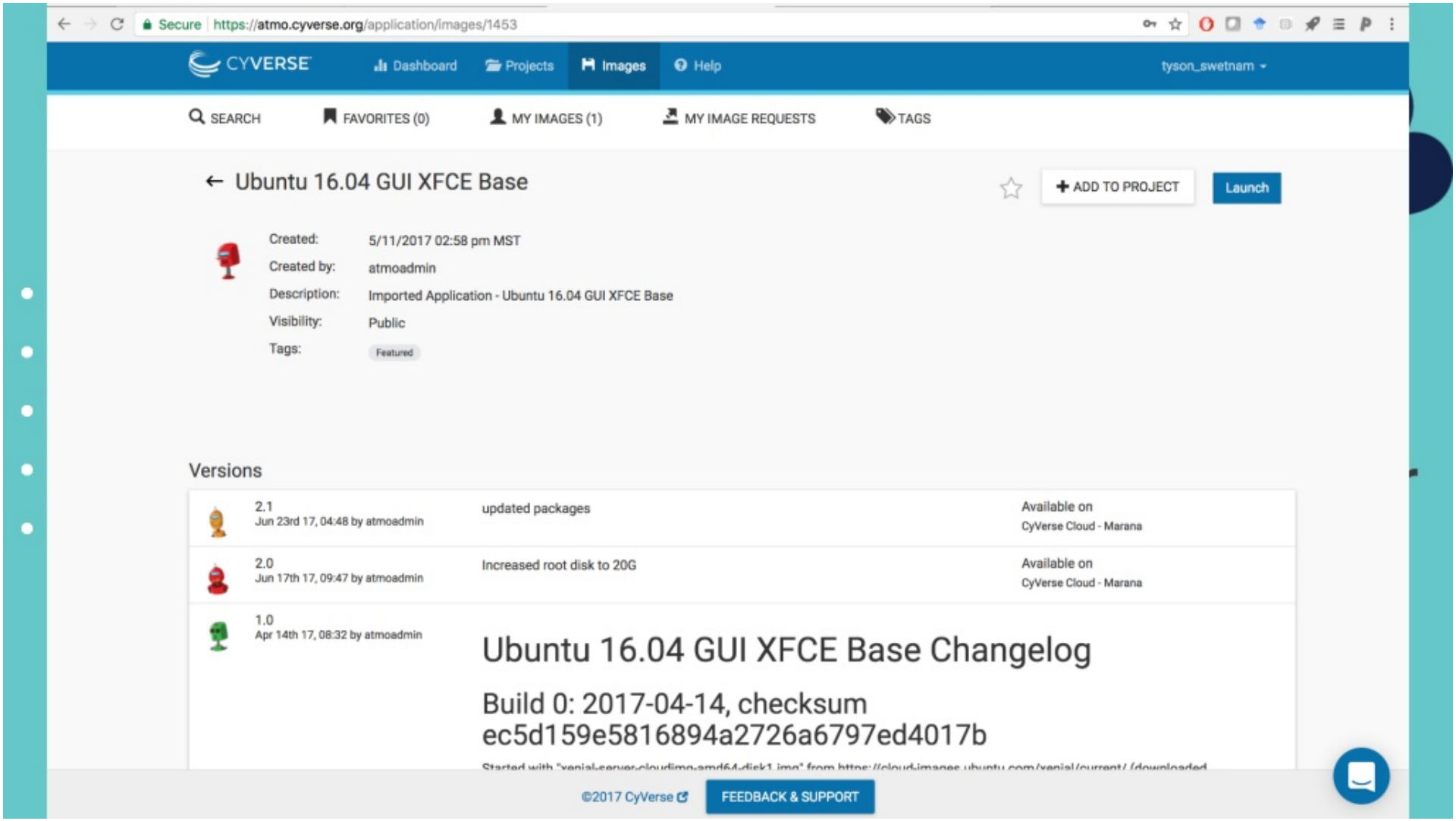
Atmosphere

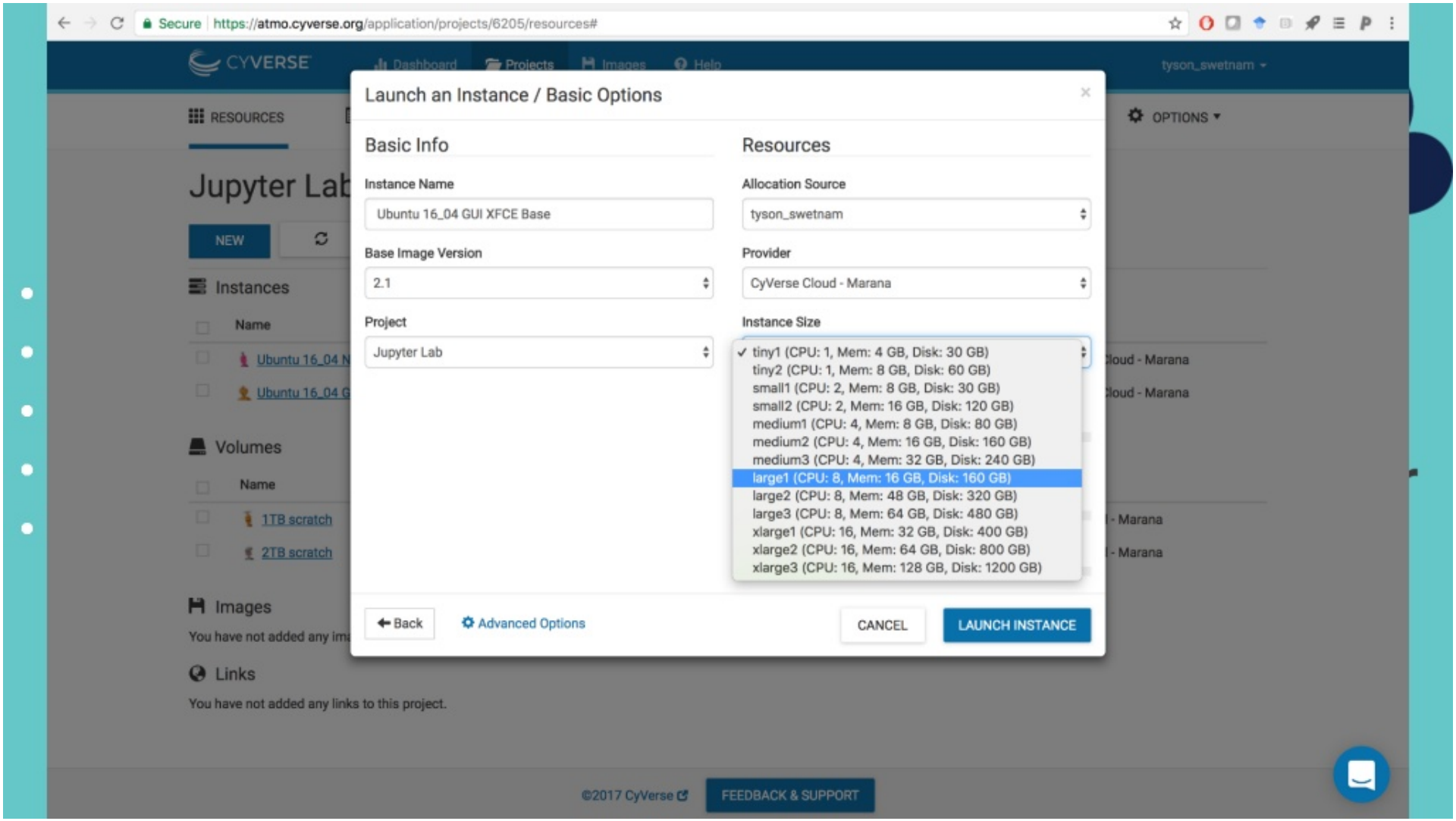
Deploy your favorite
Data Science Tools

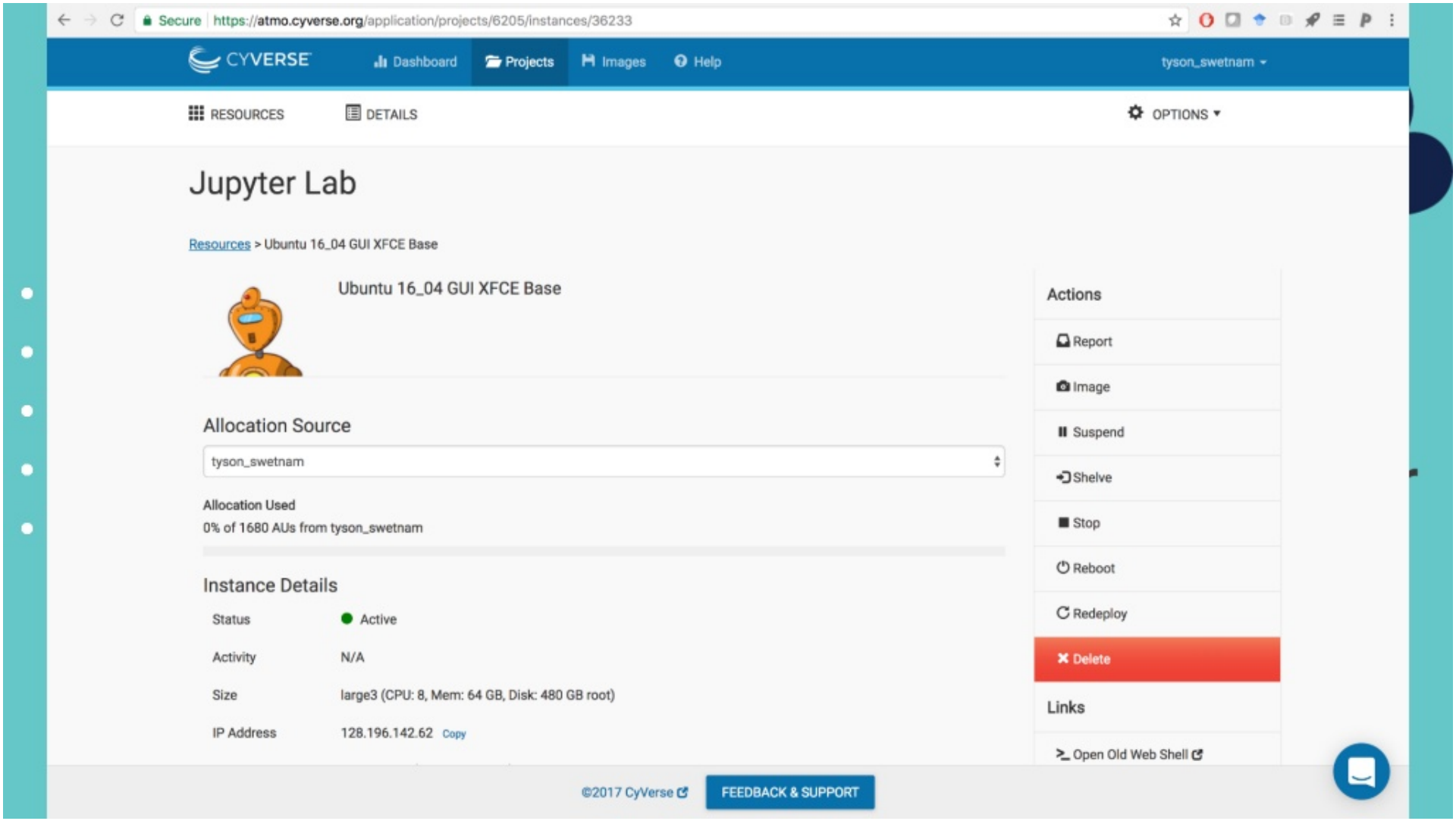
- On demand
- Linux environment
- Collaborate
- Publish images
- Easy to use interface

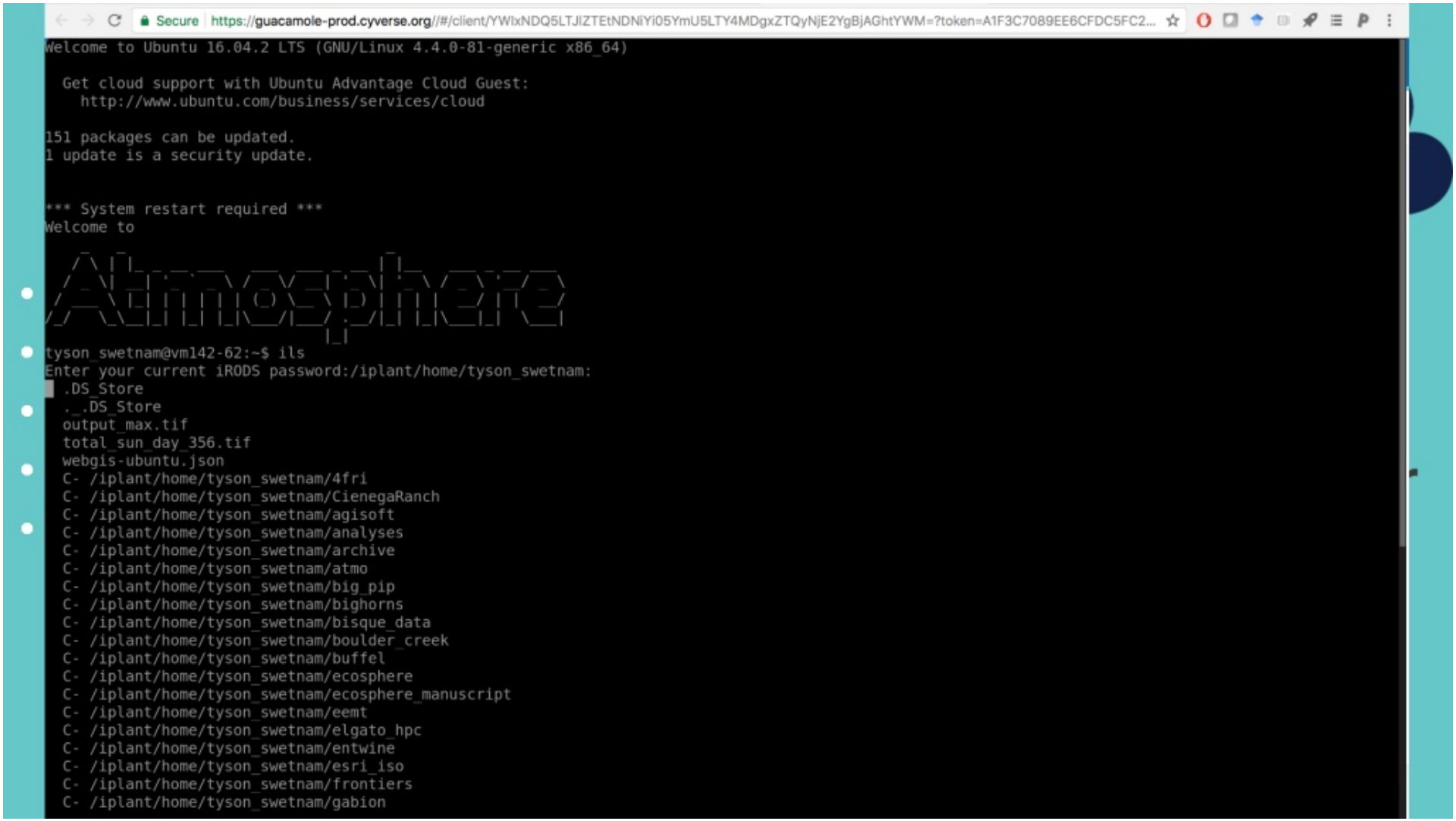
docker

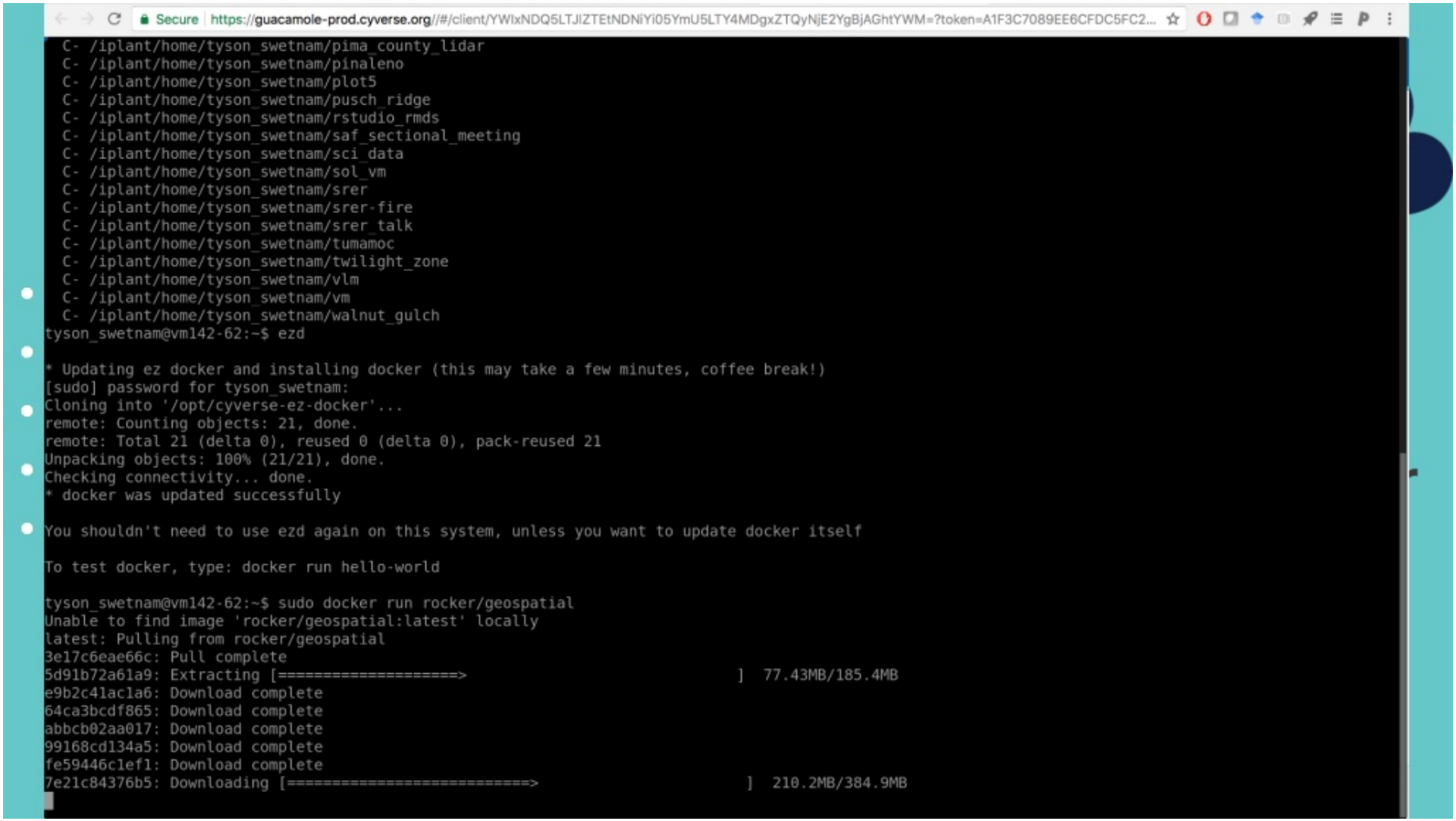

Secure | https://atmo.cyverse.org/application/dashboard
CYVERSE
Dashboard
Projects
Images
Help
tyson_swetnam
Getting Started
Launch New Instance
Browse Atmosphere's list of available images and select one to launch a new instance.
Browse Help Resources
View a video tutorial, read the how-to guides, or email the Atmosphere support team.
Change Your Settings
Modify your account settings, view your resource quota, or request more resources.
Resources Used
NEED MORE ?
Allocation Source
tyson_swetnam
Allocation
0%
0
10
20
30
40
50
60
70
80
90
100
Percent of Allocation Used
2 Instances
suspended
Provider Resources
iPlant Cloud - Tucson
CyVerse Cloud - Marana
3 Volumes
©2017 CyVerse
FEEDBACK & SUPPORT

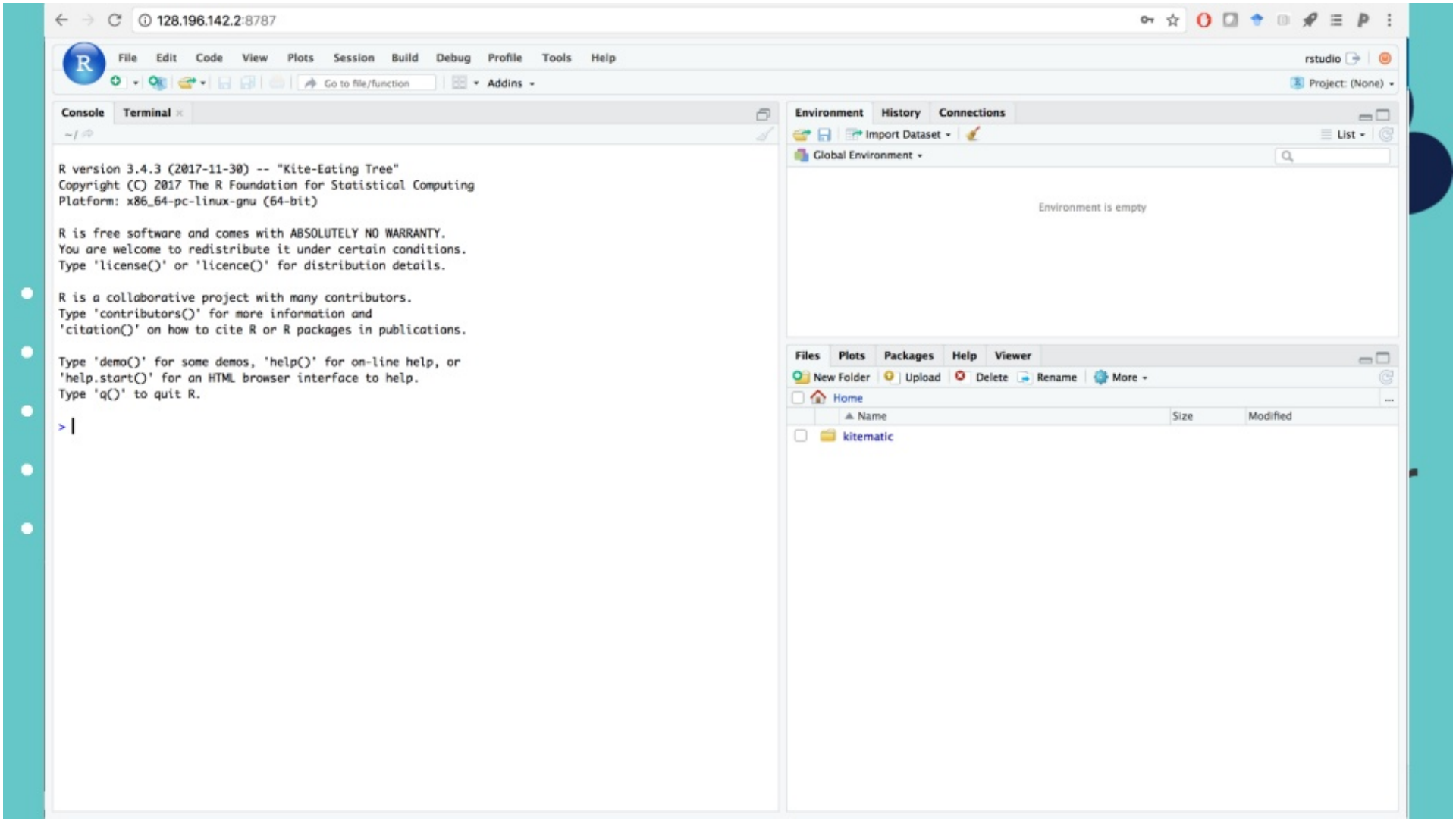

Secure https://atmo.cyverse.org/application/images/1453
CYVERSE
Dashboard
Projects
Images
Help
tyson_swetnam
SEARCH
FAVORITES (0)
MY IMAGES (1)
MY IMAGE REQUESTS
TAGS
Ubuntu 16.04 GUI XFCE Base
ADD TO PROJECT
Launch
Created: 5/11/2017 02:58 pm MST
Created by: atmoadmin
Description: Imported Application - Ubuntu 16.04 GUI XFCE Base
Visibility: Public
Tags: Featured
Versions
2.1
Jun 23rd 17, 04:48 by atmoadmin
updated packages
Available on
CyVerse Cloud - Marana
2.0
Jun 17th 17, 09:47 by atmoadmin
Increased root disk to 20G
Available on
CyVerse Cloud - Marana
1.0
Apr 14th 17, 08:32 by atmoadmin
Ubuntu 16.04 GUI XFCE Base Changelog
Build 0: 2017-04-14, checksum
ec5d159e5816894a2726a6797ed4017b
©2017 CyVerse
FEEDBACK & SUPPORT


Secure https://atmo.cyverse.org/application/projects/6205/resources#
CYVERSE
Dashboard
Projects
Images
Help
tyson_swetnam
RESOURCES
OPTIONS
Jupyter Lab
NEW
Instances
Name
Ubuntu 16_04 N
Ubuntu 16_04 G
Volumes
Name
1TB scratch
2TB scratch
Images
You have not added any ima
Links
You have not added any links to this project.
Launch an Instance / Basic Options
Basic Info
Instance Name
Ubuntu 16_04 GUI XFCE Base
Base Image Version
2.1
Project
Jupyter Lab
Resources
Allocation Source
tyson_swetnam
Provider
CyVerse Cloud - Marana
Instance Size
tiny1 (CPU: 1, Mem: 4 GB, Disk: 30 GB)
tiny2 (CPU: 1, Mem: 8 GB, Disk: 60 GB)
small1 (CPU: 2, Mem: 8 GB, Disk: 30 GB)
small2 (CPU: 2, Mem: 16 GB, Disk: 120 GB)
medium1 (CPU: 4, Mem: 8 GB, Disk: 80 GB)
medium2 (CPU: 4, Mem: 16 GB, Disk: 160 GB)
medium3 (CPU: 4, Mem: 32 GB, Disk: 240 GB)
large1 (CPU: 8, Mem: 16 GB, Disk: 160 GB)
large2 (CPU: 8, Mem: 48 GB, Disk: 320 GB)
large3 (CPU: 8, Mem: 64 GB, Disk: 480 GB)
xlarge1 (CPU: 16, Mem: 32 GB, Disk: 400 GB)
xlarge2 (CPU: 16, Mem: 64 GB, Disk: 800 GB)
xlarge3 (CPU: 16, Mem: 128 GB, Disk: 1200 GB)
Back
Advanced Options
CANCEL
LAUNCH INSTANCE
©2017 CyVerse
FEEDBACK & SUPPORT


Secure https://atmo.cyverse.org/application/projects/6205/instances/36233
CYVERSE
Dashboard
Projects
Images
Help
tyson_swetnam
RESOURCES
DETAILS
OPTIONS
Jupyter Lab
Resources > Ubuntu 16_04 GUI XFCE Base
Ubuntu 16_04 GUI XFCE Base
Allocation Source
tyson_swetnam
Allocation Used
0% of 1680 AUs from tyson_swetnam
Instance Details
Status
Active
Activity
N/A
Size
large3 (CPU: 8, Mem: 64 GB, Disk: 480 GB root)
IP Address
128.196.142.62
Copy
Actions
Report
Image
Suspend
Shelve
Stop
Reboot
Redeploy
Delete
Links
Open Old Web Shell
©2017 CyVerse
FEEDBACK & SUPPORT


Secure https://guacamole-prod.cyverse.org//#/client/YWlxNDQ5LTJlZTEtNDNiYi05YmU5LTY4MDgxZTQyNjE2YgBjAGhtYWM=?token=A1F3C7089EE6CFDC5FC2...
Welcome to Ubuntu 16.04.2 LTS (GNU/Linux 4.4.0-81-generic x86_64)
Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud
151 packages can be updated.
1 update is a security update.
*** System restart required ***
Welcome to
tyson_swetnam@vm142-62:~$ ils
Enter your current iRODS password:/iplant/home/tyson_swetnam:
.DS_Store
._.DS_Store
output_max.tif
total_sun_day_356.tif
webgis-ubuntu.json
C- /iplant/home/tyson_swetnam/4fri
C- /iplant/home/tyson_swetnam/CienegaRanch
C- /iplant/home/tyson_swetnam/agisoft
C- /iplant/home/tyson_swetnam/analyses
C- /iplant/home/tyson_swetnam/archive
C- /iplant/home/tyson_swetnam/atmo
C- /iplant/home/tyson_swetnam/big_pip
C- /iplant/home/tyson_swetnam/bighorns
C- /iplant/home/tyson_swetnam/bisque_data
C- /iplant/home/tyson_swetnam/boulder_creek
C- /iplant/home/tyson_swetnam/buffel
C- /iplant/home/tyson_swetnam/ecosphere
C- /iplant/home/tyson_swetnam/ecosphere_manuscript
C- /iplant/home/tyson_swetnam/eemt
C- /iplant/home/tyson_swetnam/elgato_hpc
C- /iplant/home/tyson_swetnam/entwine
C- /iplant/home/tyson_swetnam/esri_iso
C- /iplant/home/tyson_swetnam/frontiers
C- /iplant/home/tyson_swetnam/gabion


Secure https://guacamole-prod.cyverse.org//#/client/YWIxNDQ5LTJlZTEtNDNiYi05YmU5LTY4MDgxZTQyNjE2YgBjAGhtYWM=?token=A1F3C7089EE6CFDC5FC2...
C- /iplant/home/tyson_swetnam/pima_county_lidar
C- /iplant/home/tyson_swetnam/pinaleno
C- /iplant/home/tyson_swetnam/plot5
C- /iplant/home/tyson_swetnam/pusch_ridge
C- /iplant/home/tyson_swetnam/rstudio_rmds
C- /iplant/home/tyson_swetnam/saf_sectional_meeting
C- /iplant/home/tyson_swetnam/sci_data
C- /iplant/home/tyson_swetnam/sol_vm
C- /iplant/home/tyson_swetnam/srer
C- /iplant/home/tyson_swetnam/srer-fire
C- /iplant/home/tyson_swetnam/srer_talk
C- /iplant/home/tyson_swetnam/tumamoc
C- /iplant/home/tyson_swetnam/twilight_zone
C- /iplant/home/tyson_swetnam/vlm
C- /iplant/home/tyson_swetnam/vm
C- /iplant/home/tyson_swetnam/walnut_gulch
tyson_swetnam@vm142-62:~$ ezd
* Updating ez docker and installing docker (this may take a few minutes, coffee break!)
[sudo] password for tyson_swetnam:
Cloning into '/opt/cyverse-ez-docker'...
remote: Counting objects: 21, done.
remote: Total 21 (delta 0), reused 0 (delta 0), pack-reused 21
Unpacking objects: 100% (21/21), done.
Checking connectivity... done.
* docker was updated successfully
You shouldn't need to use ezd again on this system, unless you want to update docker itself
To test docker, type: docker run hello-world
tyson_swetnam@vm142-62:~$ sudo docker run rocker/geospatial
Unable to find image 'rocker/geospatial:latest' locally
latest: Pulling from rocker/geospatial
3e17c6eae66c: Pull complete
5d91b72a61a9: Extracting [==========>] 77.43MB/185.4MB
e9b2c41ac1a6: Download complete
64ca3bcdf865: Download complete
abbcb02aa017: Download complete
99168cd134a5: Download complete
fe59446c1ef1: Download complete
7e21c84376b5: Downloading [==========>] 210.2MB/384.9MB


128.196.142.2:8787
File Edit Code View Plots Session Build Debug Profile Tools Help
rstudio
Go to file/function
Addins
Project: (None)
Console
Terminal
R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
kitematic

Atmosphere

Deploy your favorite
Data Science Tools

- On demand
- Linux environment
- Collaborate
- Publish images
- Easy to use interface

Atmosphere

Deploy your favorite
Data Science Tools

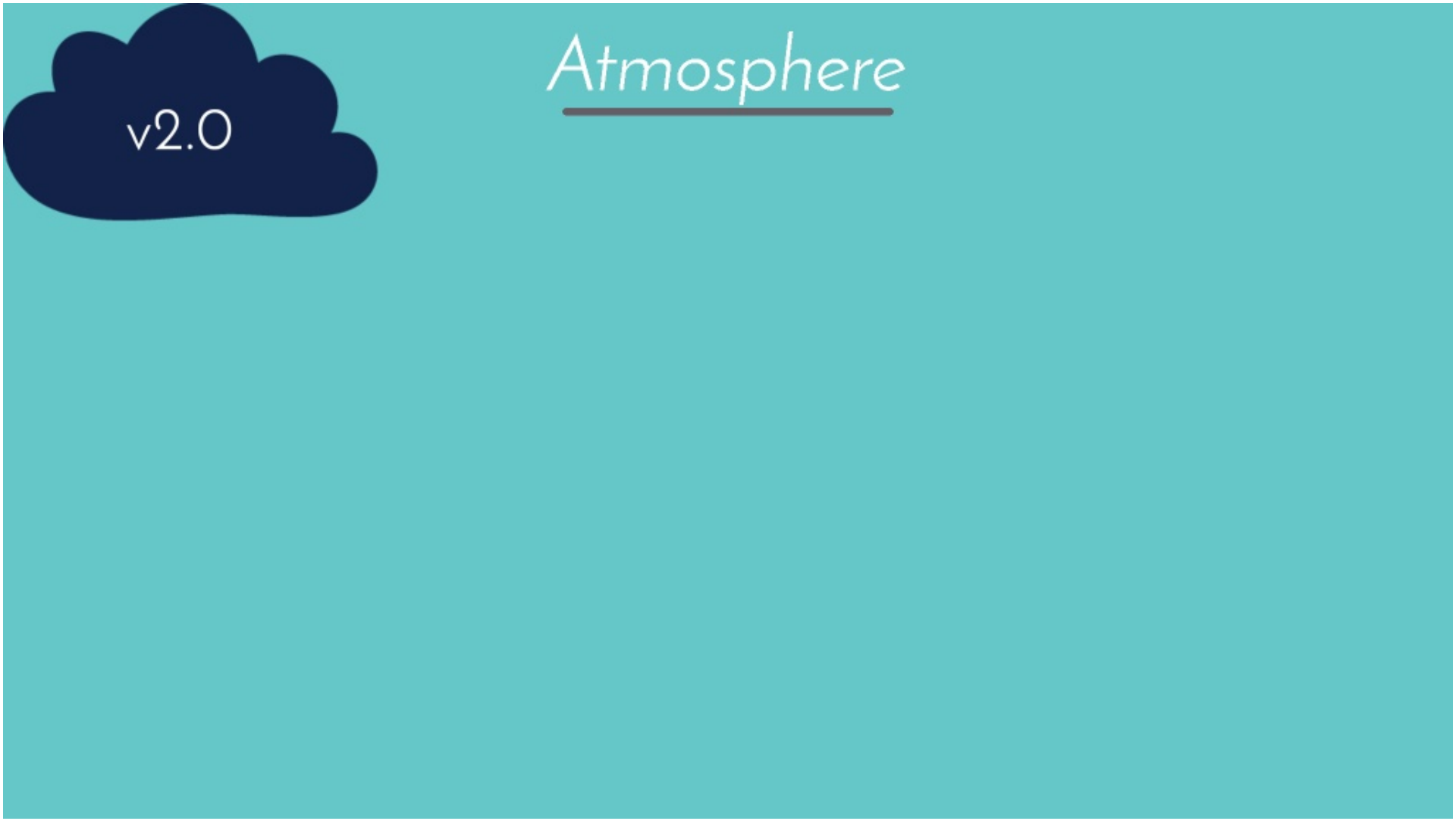
- On demand
- Linux environment
- Collaborate
- Publish images
- Easy to use interface

RStudio®

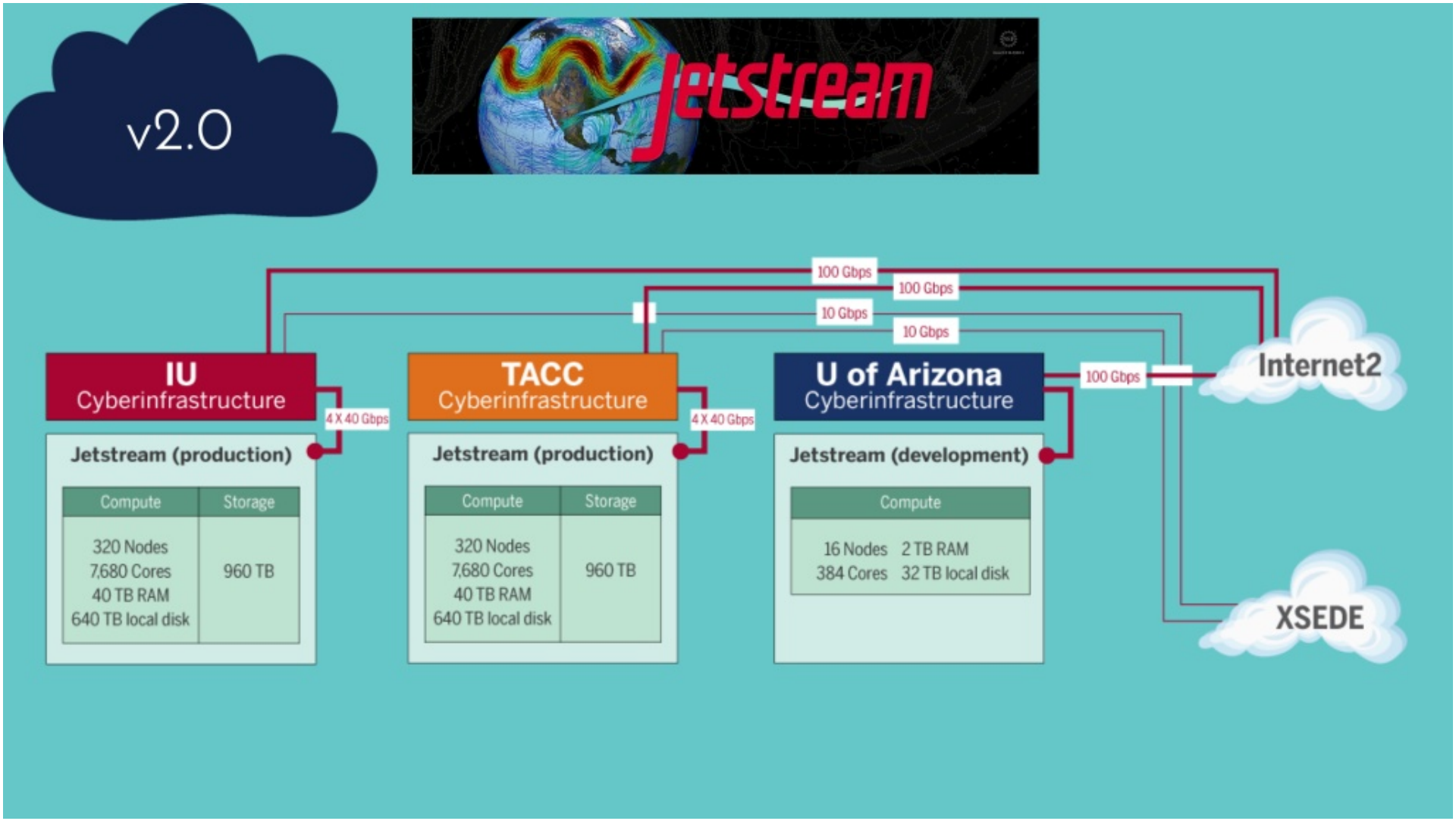
docker

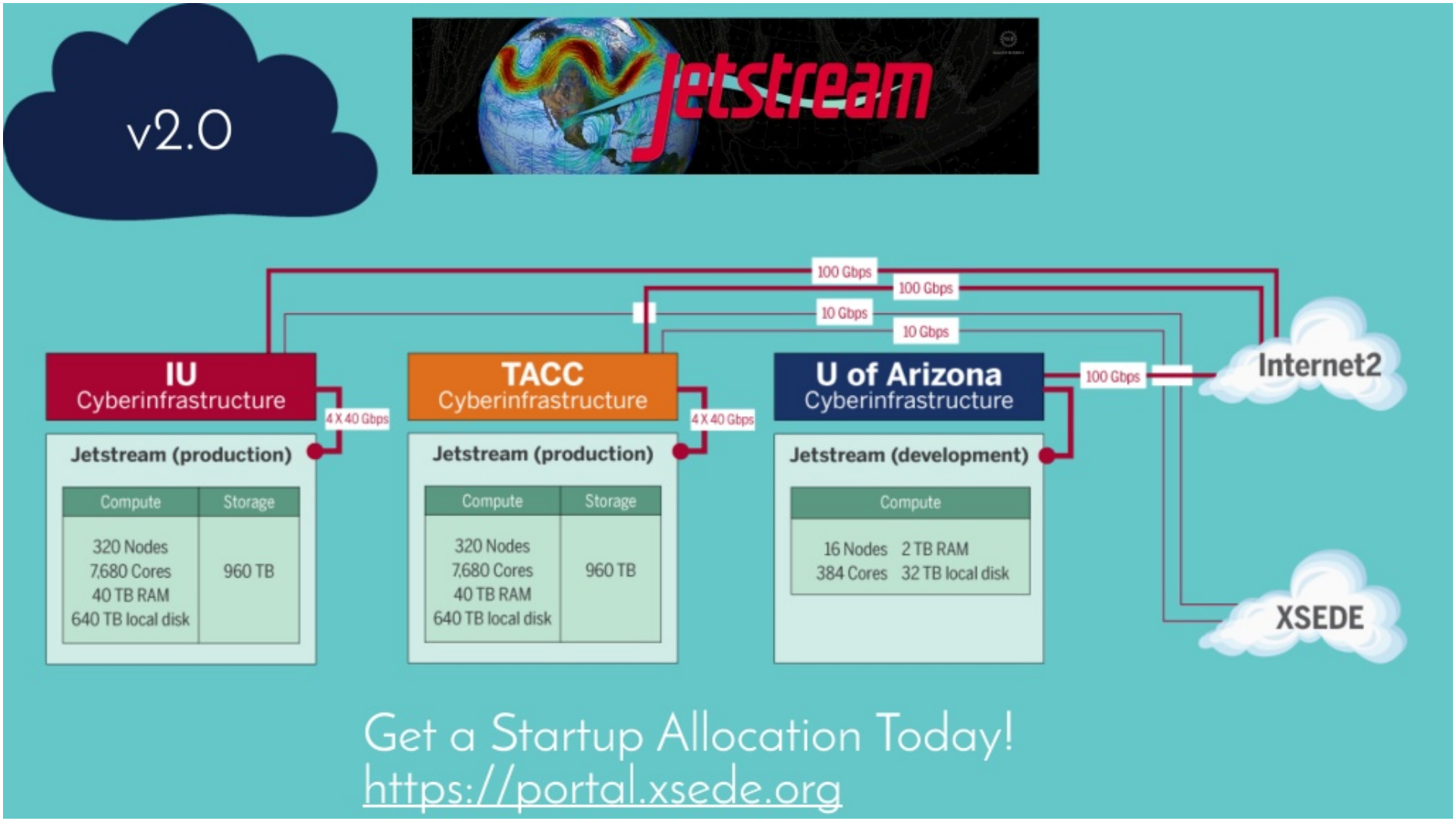
jupyter

Very Popular → Oversubscribed

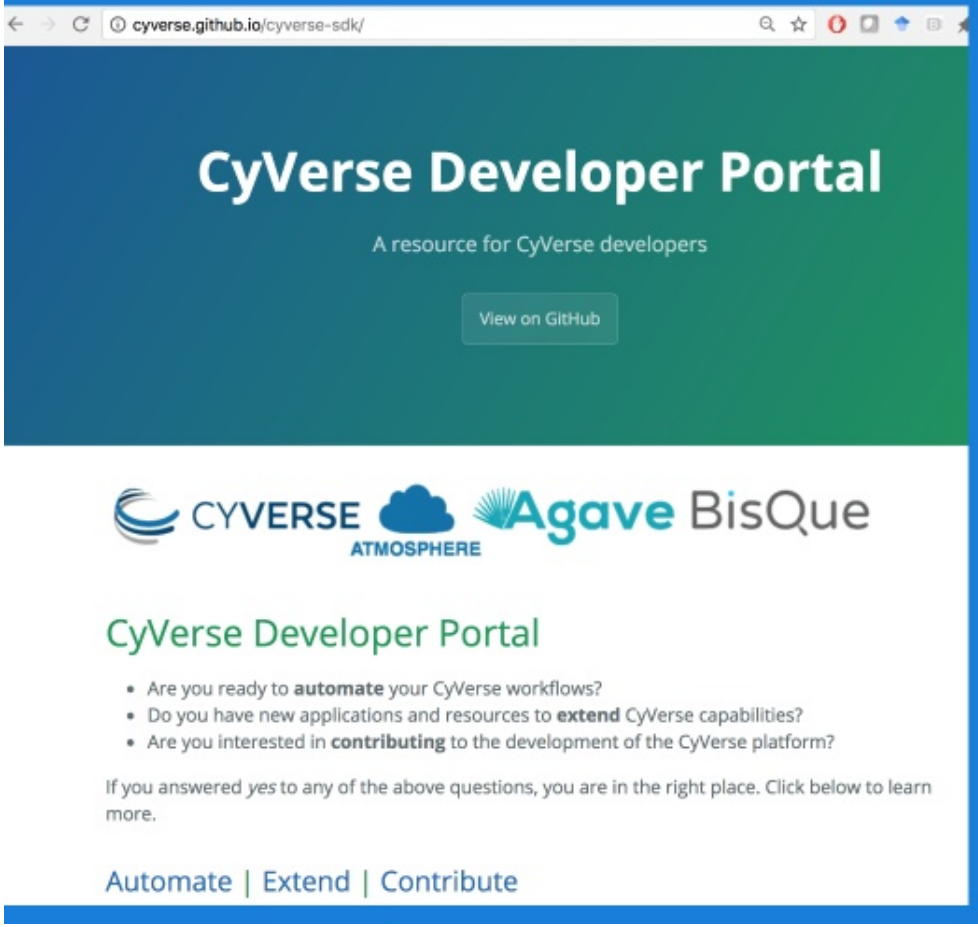

v2.0
Atmosphere

v2.0


v2.0
Jetstream
100 Gbps
100 Gbps
10 Gbps
10 Gbps
IU
Cyberinfrastructure
4 X 40 Gbps
Jetstream (production)
Compute
Storage
320 Nodes
7,680 Cores
40 TB RAM
640 TB local disk
960 TB
TACC
Cyberinfrastructure
4 X 40 Gbps
Jetstream (production)
Compute
Storage
320 Nodes
7,680 Cores
40 TB RAM
640 TB local disk
960 TB
U of Arizona
Cyberinfrastructure
Jetstream (development)
Compute
16 Nodes 2 TB RAM
384 Cores 32 TB local disk
100 Gbps
Internet2
XSEDE


v2.0
Jetstream
100 Gbps
100 Gbps
10 Gbps
10 Gbps
IU
Cyberinfrastructure
4 X 40 Gbps
Jetstream (production)
Compute
Storage
320 Nodes
7,680 Cores
40 TB RAM
640 TB local disk
960 TB
TACC
Cyberinfrastructure
4 X 40 Gbps
Jetstream (production)
Compute
Storage
320 Nodes
7,680 Cores
40 TB RAM
640 TB local disk
960 TB
U of Arizona
Cyberinfrastructure
Jetstream (development)
Compute
16 Nodes 2 TB RAM
384 Cores 32 TB local disk
100 Gbps
Internet2
XSEDE
Get a Startup Allocation Today!
https://portal.xsede.org

Science APIs



1,218 Apps
(96 contributors)

72,874 analysis
(10+ HPC resources)

Lessons learned in cyberinfrastructure management and data hosting

- For reproducability, data must keep its provenance and attribution throughout life cycle as part of the Research Object
- Make developers and informaticians available to answer researchers questions is key and ensures researchers get the help they need in a timely manner.
- Establish where (re-)analyses will take place. Make data reuse easy.

Lessons learned in cyberinfrastructure management and data hosting

- For reproducability, data must keep its provenance and attribution throughout life cycle as part of the Research Object

- Make developers and informaticians available to answer researchers questions is key and ensures researchers get the help they need in a timely manner.

- Establish where (re-)analyses will take place. Make data reuse easy.

Lessons learned in cyberinfrastructure management and data hosting

- For reproducability, data must keep its provenance and attribution throughout life cycle as part of the Research Object

- Make developers and informaticians available to answer researchers questions is key and ensures researchers get the help they need in a timely manner.

- Establish where (re-)analyses will take place. Make data reuse easy.

Lessons learned in cyberinfrastructure management and data hosting

- For reproducability, data must keep its provenance and attribution throughout life cycle as part of the Research Object

- Make developers and informaticians available to answer researchers questions is key and ensures researchers get the help they need in a timely manner.

- Establish where (re-)analyses will take place. Make data reuse easy.

Lessons learned in cyberinfrastructure management and data hosting

- For reproducability, data must keep its provenance and attribution throughout life cycle as part of the Research Object

- Make developers and informaticians available to answer researchers questions is key and ensures researchers get the help they need in a timely manner.

- Establish where (re-)analyses will take place. Make data reuse easy.

Acknowledgements

- CyVerse is funded by NSF DBI-0735191 and DBI-1265383
- Jetstream is funded by NSF ACI-1445604
- Agave is funded by NSF ACI-1450459, ACI-1127210, DBI-0735191, IOS-1237931, IOS-1237931, DBI-1262414, and 1R01A1097403.

https://cyverse.org

My Contact Info:

tswetnam@cyverse.org

@tswetnam

tyson-swetnam


CyVerse Data Commons: lessons learned in cyberinfrastructure management and data hosting from the Life Sciences
AGU Fall Meeting December 11, 2017 Sesson IN12B-07
Tyson L. Swetnam, Ramona L. Walls, & Nirav Merchant
BIO5 Institute, University of Arizona, Tucson AZ
Enter Here
CYVERSE